Table of Contents

	Page
Introduction	1
Applications	
Circuit Construction	
Chip Installation	3
Non-Volatile Memory Backup	
Using the Speech Recognition Circuit	
Training	4
To Train	
Testing Recognition	
Error Codes	
Clearing Memory	
Robust & Simulated Independent Recognition	5
Homonyms	6
The Voice Stress & Excitement	
Interfacing the Circuit to the Outside World	
Error Codes	7
Interface Circuit	
Voice Security System	8
CPU Mode	
AUI Aural User Interface	
Learning To Listen	9
Speaker Dependent / Speaker Independent	
Recognition Style	10
More on the HM2007	
Parts List	11

Introduction

The speech recognition kit is a complete easy to build programmable speech recognition circuit. Programmable, in the sense that you program the words (or vocal utterances) you want the circuit to recognize. This kit allows you to experiment with many facets of speech recognition technology.

Some features of the kit follows:

Self-Contained stand alone speech recognition circuit
User Programmable
40 word vocabulary
Multi-lingual
Non-volatile memory back up
Easily interfaced to control external circuits & appliances

Speech Recognition will become the method of choice for controlling appliances, toys, tools and computers.

At its most basic level, speech controlled appliances and tools allow the user to perform parallel tasks, (i.e. hands and eyes are busy elsewhere) while working with the tool or appliance.

The heart of the circuit is the HM2007 speech recognition IC. The IC can recognize 40 words, each word with a maximum length of .96 seconds. HM2007 chips can be cascaded to provide a larger word recognition library.

Applications

There are several major areas for application of voice recognition technology.

Speech controlled Appliances and Toys.
Speech Assisted Computer Games
Speech Assisted Virtual Reality
Telephone Assistance Systems.
Data Entry
Voice Recognition Security.
Speech to Speech Translation.

Circuit Construction

The schematic is shown in figure 1. The components ar mounted on the top side of the board and are soldered on ti foil side of the PCB board.

Begin construction by soldering the two 16 dip 220 ohm resistors paks. Next solder the integrated circuit sockets to the PC board. The seven segmented display chips use 14 pin sockets. The 7448 chips use 16 pin sockets. The 74LS373 uses a 20 pin socket. The HM2007 uses the 48 pin socket and the 8K static ram uses a 28 pin socket.

Solder resistor R1, it has a nominal value of 22K. Its color bands are red, red, orange, and gold.

Solder resistor R2, it has a nominal value of 6.8K. Its color bands are blue, gray, red and gold.

Solder resistor R3, it has a nominal value of 100 K. Mount and solder diodes D1 and D2. Make sure the black band faces the correct direction as shown on the drawings.

Next solder jumper wire J1.

Now solder the 3.57 MHz crystal. It is identified as XTAL on the parts placement drawing. Solder the red LED next. The long lead of the LED is positive. Next solder all the capacitors C1 to C3.

Next solder the 7805 voltage regulator and switch. Finish up the board by soldering in the microphone and battery cap. In both cases the red wire is the positive lead.

The keypad is connected by seven wires to the circuit board. Solder seven (7) wires from the keypad to the circuit, see figures 2 and 3. Pay particular attention in wiring the keypad, the wires don't always follow in sequence.

Chip Installation

Begin installing the integrated circuit (IC) chips by first identifying the top of the chip. The top of the chip has an marker, a half circle cutout. Orientated the IC chips exactly as shown in the schematic and placed them into their sockets.

Non-Volatile Memory Back-up

With the memory backup in place, the word patterns are retained in memory when the circuit is turned off. This alleviates having to retrain the circuit whenever it is turned off.

The circuit has a 3V DC input leads to make the static ram used on the board non-volatile. 3VDC can be supplied from two "AA" batteries. The current drain on the batteries is so low, that the lifetime of the batteries will approach their shelf life.

The circuit is compatible with Dallas Non-volatile SRAM memory chips (DS1225D-85). Since these SRAM chips have a built in battery no external 3VDC power is required.

Using The Speech Recognition Circuit

The keypad and digital display are used to communicate with and program the HM2007 chip.

Keypad:

The keypad is made up of 12 keys.

1	2	3
4	5	6
7	8	9
*	0	#
Clear		Train

Note that the "*" key equals Clear and the "#" key equals Train.

When the circuit is turned on the board displays "00" on the digital display, lights the red LED (READY) and waits for a command.

Training

To Train:

Press "01" (display will show "01"), then press "#" to train word. (Training).

Speak the target word into the microphone. If the circuit accepted the input the LED will flash momentarily. The word spoken (entered) is now identified as the "01" word. If the LED did not flash start over by pressing "O1" then "T".

You may continue training new words in the circuit. Press "02" then "#" to train the second word and so on. The circuit will accept and recognize up to forty words (numbers 1 thru 40). It is not necessary to train all word spaces. If you only require 10 target words that's all you need to train.

Testing Recognition:

Repeat a trained word into the microphone. The number of the word should be displayed on the segmented display. For instance if the word "directory" was trained as word number 25. Saying the word "directory" into the microphone will cause the number 25 to be displayed.

Error Codes:

The chip provides the following error codes.

55 = too long
66 = too short
77 = no match

Clearing Memory:

To erase SRAM memory press "99" then "C". The numbers will quickly scroll by in the digital display as the memory is erased.

Robust & Simulated Independent Recognition

The speech recognition system is speaker dependent. Meaning that the voice that trained the system has the highest recognition accuracy. You can simulate independent speech recognition.

To make the system more robust and at the same time simulate speaker i ndependence one uses more than one word space for each target word. The best system we found uses four word spaces per target word. Therefore you can have four different enunciations of the target word.

The four word spaces are chosen to minimize software and hardware interfaces into the circuit. To accomplish this the four word spaces are chosen so that they all have the same Lest Significant Digit (LSD). Doing this the words can be recognized by just decoding the lest significant digit (number) on the digital display.

Using this system word spaces; 01, 11, 21 and 31 are

allocated to the first target word. The Most Significant Digit (MSD) is dropped. By only decoding the lest significant digit number, in this case 1 of "X1" (where X is any number) we can recognize the target word.

We do this for the remaining word spaces. For instances, the second target word will use word spaces 02, 12, 22 and 32. We continue in this manner until all the words are programmed.

If you are experimenting with speaker independence use different people when training a target word. This will enable the system to recognize different voices, inflections and enunciations of the target word. The more system resources that are allocated for independent recognition the more robust the circuit will become.

If you are experimenting with designing the most robust and accurate system possible, train target words using one voice with different inflections and enunciations of the target word.

Homonyms

Homonyms are words that sound alike. For instance the words; cat, bat, sat and fat sound alike. Because of their like sounding nature they can confuse the speech recognition circuit. When choosing target words for your system steer clear of homonyms.

The Voice with Stress & Excitement

Stress and excitement alters ones voice. This affects the accuracy of the circuit's recognition. For instances assume you are sitting at your workbench and you program the target words like, fire, left, right, forward, etc., into the circuit. Then you use the circuit to control a flight simulator game or Doom or Duke Nukem. Well when you're playing the game you likely be yelling FIRE!. Fire!.. FIRE!!. LEFT..., go RIGHT!. In the heat of the action you're voice will appear much different than when your were sitting down relaxed and programming the circuit. To achieve a higher accuracy word recognition accuracy one needs to mimic the excitement in ones voice when programming the circuit.

These factors should be kept in mind to achieve the high accuracy possible from the circuit. This becomes increasingly important when the speech recognition is taken out of the lab and put to work in the outside world.

Interfacing the Circuit to the Outside World

The circuit design for interfacing the speech recognition system to the outside controls 10 switches. This fits in well with the robust speech recognition system discussed previously. While the effective vocabulary drops from forty words to ten words, we gain in a more robust and accurate system.

Error Codes

The decoding circuit must recognizes the word numbers from error codes. So the circuit must be designed to recognize error codes 55, 66 and 77 and not confuse them with word spaces 5,6 and 7. This is accomplished using an OR gate and a NAND gate connected to the MSB. Whenever an error code is generated, the word number is ignored.

Interface Circuit

Figure 3 is the schematic for the interface circuit. The circuit connects to the 9 pin interface header on the circuit board.

The 4028 has ten output lines. Whatever number is displayed on the LSD the corresponding line number off the 4028 will be brought high. The high signal from the 4028 can be connected to a NPN transistor to control a DC Load as shown in box A, or control an AC or DC load using a simple relay as shown in box B.

The disadvantage to using a simple set up like this is that only one switch out of 10 may be turned on at any given time. This doesn't make for a very good system. So a solution must be found that allows one to turn on or off any line without changing the status of any other line.

This can be implemented by inserting a flip-flop, shown in box C, inbetween the 4028 and the NPN transistor. The 4013 IC contains two flip flops, only one is shown in the box C drawing. The flip flop acts like simple memory. When the input line is brought high, its output line goes high. Turning on the NPN transistor. When the input line is brought low, the output line still stays high. When the flip flop receives a second high signal on its input line it brings the output low.

OK lets see how this works in the real world. Let assume you have the power to a printer connected through the speech recognition circuit and a 4013 controlled switch or relay. The target word is "printer". When you want to turn the printer on you use the command word "printer". The circuit recognizes the word and applies power to the printer. At this point you can also turn on (or off) any other circuit connected to the speech board, because when the high signal goes low the 4013 keeps its output high. When you are ready to turn off the power to the printer all you have to do is repeat the command "printer". The second time the line is brought high, the output of the 4013 is brought low.

The same command is used to turn the unit on and off. Any of the line can be turned on or off without affecting the status of the other output lines

Voice Security System

This circuit isn't designed for a voice security system in a commercial application, but that should not prevent anyone from experimenting with it for that purpose. You may want to use three or four keywords that must be spoken and recognized in sequence in order to open a lock or allow entry. Bandpass filtering the audio input should make the system more reliable.

CPU Mode

The HM2007 speech recognition chip has a CPU mode to be used when connected to a host computer system. Connecting the chip to the IBM PC bus isn't a problem. It will require driver software to control training, storing, and identify recognition. When the circuit is interfaced to a computer it is possible to store trained words in a program file that can be upload into the circuit's static ram when started. For anyone wanting to interface the HM2007 to a PC bus I recommend purchasing the DS-HM2007 data booklet.

AUI "Aural User Interface" Technology

Most speech recognition systems require the use of a host personal computer. These are speech recognition programs that run continuously in the background of the computers operating system. Most programs require the computer be equipped with a compatible soundcard for speech input.

Companies are actively developing Aural User Interfaces (AUIs). The AUI design determines how a computers (or circuit) listens for commands, confirm (optional) and carries out that command.

Mixing visual and aural information has been found not to be effective. When products tried to visually confirm an aural command it quickly became apparent that mixing aural and visual information reduced efficiency. Products are needed that can understand (recognize) commands given in an unstructured and efficient methods in which people typically communicate verbally. To this end manufactures, designers and programmers should strive for effective AUI's that does not require secondary visual checking of input.

Learning to Listen

We take our speech recognition abilities for granted. Listening to one person speak among several at a party is beyond the capabilities of today's speech recognition systems. Speech recognition systems have a hard time separating and filtering out what should be considered extraneous noise.

Speech recognition is not speech understanding. Understanding the meaning of words is a higher intellectual function. Because a computer can respond to a vocal command does not mean it understands the command spoken. Future voice recognition system may have the ability to distinguish nuances and meaning of words, to "Do what I mean, not what I say!"

Speaker Dependent / Speaker Independent

Speech recognition is divided into two broad processing categories, speaker dependent and speaker independent.

Speaker dependent systems are trained by the individual who will be using the system. These system are capable of achieving a high command count and better than 95% accuracy for word recognition. The draw back to this approach is that the system responds accurately only to the individual who trained the system. This is the most common approach employed in software for personal computers.

Speaker independent is a system trained to respond to a word regardless of who speaks. Therefore the system must respond to a large variety of speech patterns, inflections and enunciations of the target word. The command word count is usually lower than the speaker dependent however high accuracy can still be maintain within processing limits. Industrial requirements more often need speaker independent voice systems, such as the AT&T system used in telephone systems.

Recognition Style

In addition to the speaker dependent/independent classification, speech recognition systems must also contend with the style of speech it can recognize. They are three style classifications of speech: isolated, connected and continuous.

Isolated. These are words that are spoken separately or isolated. This is the most common style speech recognition systems available today. The user must pause between each word or command spoken. The speech recognition circuit is set up to identify isolated words of .96 second lengths.

Connected. This is a half way point between isolated word and continuous speech recognition. Allows users to speak multiple words. The HM2007 has an optional mode to identify words or phrases 1.92 seconds in length. This reduces the word recognition dictionary number to 20.

Continuous. This is the natural conversational speech we are use to in everyday life. It is extremely difficult for a recognizer to shift through the text as the word tend to merge together. For instance, "Hi, how are you doing?" sounds like "Hi,.howyadoin" Continuous speech recognition systems are on the market and are under continual development.

More on The HM2007 CHIP:

The HM2007 is a single chip CMOS voice recognition LSI (Large Scale Integration) circuit. The chip contains a analog front end, voice analysis, recognition, and system control functions. The chip man be used in a stand alone or CPU connected.

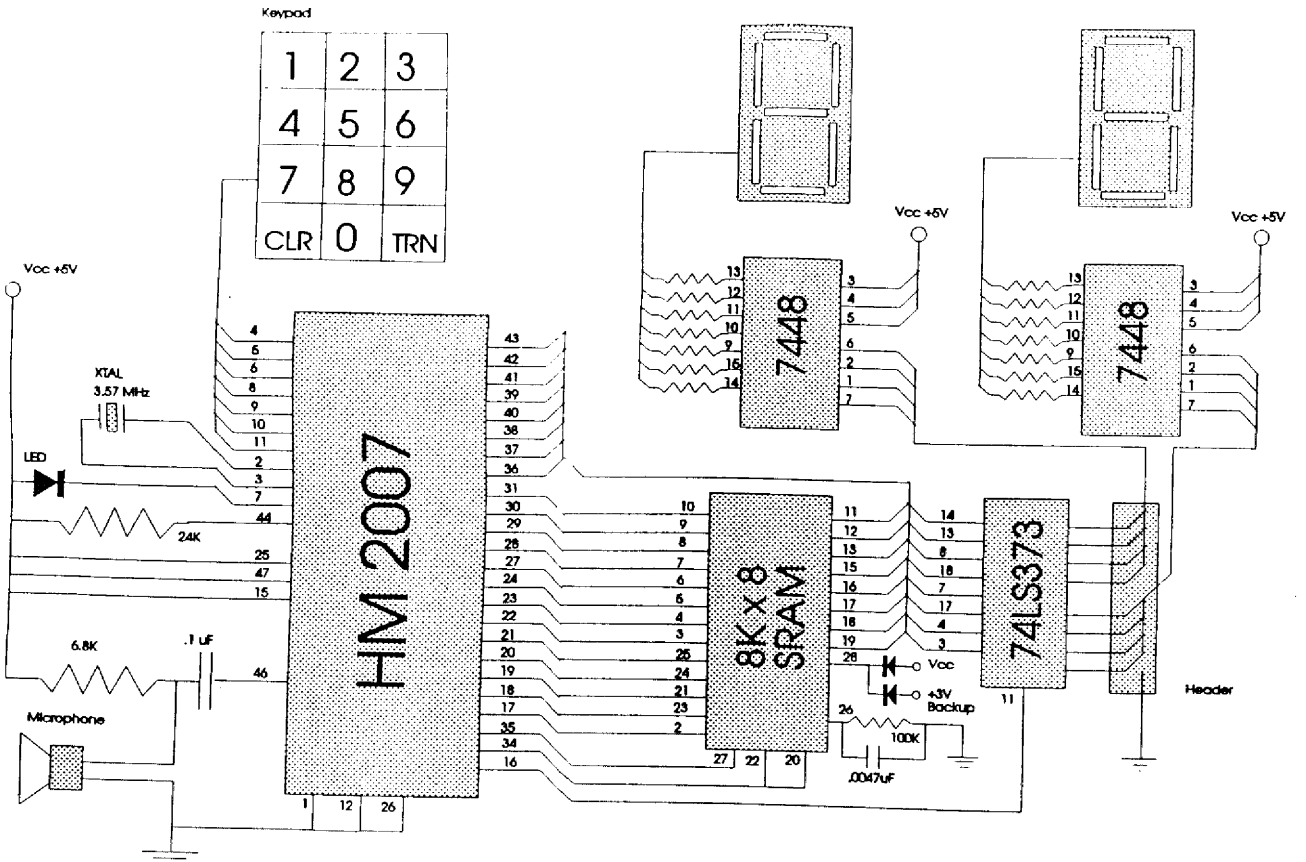
Features:

- Single chip voice recognition CMOS LSI.
- Speaker-dependent
- External RAM support
- Maximum 40 word recognition
- Maximum word length 1.92 seconds
- Microphone support
- Manual and CPU modes available
- Response time less than 300 milliseconds
- 5V power supply

More information on the HM2007 chip is available in the HM2007 data booklet. DS-HM2007

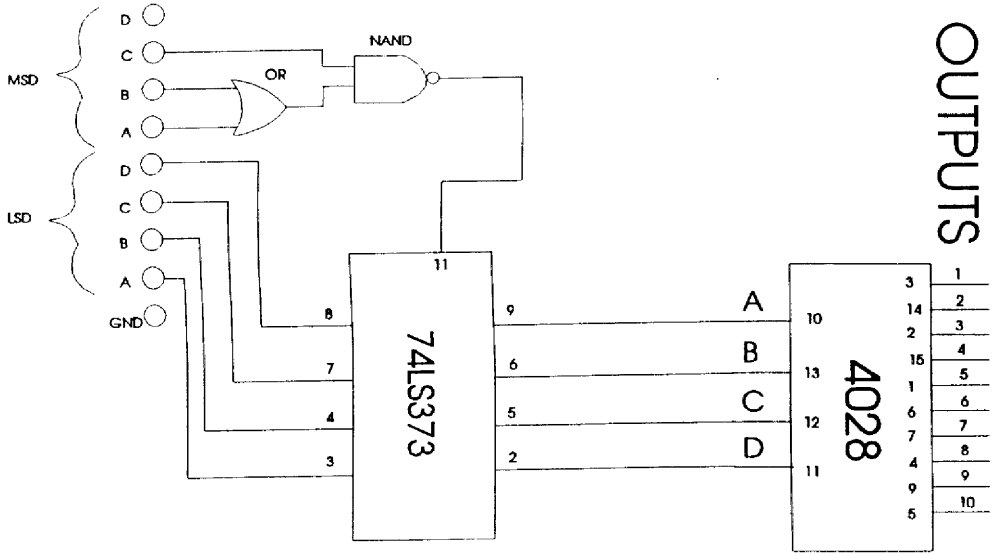
QUAN	ITEM	
1	IC1	HM2007 IC
1	IC2	SRAM 8K X 8
1	IC3	74LS373
2	IC4 & IC5	7448
2		220 ohm IC's
1		XTAL 3.57 Mhz
1		PCB
1		Keypad
1		PC mounted switch
2		7 Segment Displays
1		Microphone
1		Battery Clip
1	R1	22K 1/4 watt resistor
1	R2	5.6K 1/4 watt resistor
1	R3	100K 1/8 watt resistor
1	C1	.1 uF cap
1	C2	100 uF cap
1	C3	.0047 uF Cap.

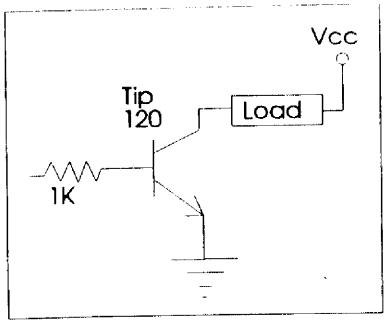
Images Company


Keypad
1 2 3
4 5 6
7 8 9
CLR 0 TRN
Vcc +5V
XTAL
3.57 MHz
LED
24K
6.8K
.1 uF
Microphone
HM 2007
7448
7448
Vcc +5V
Vcc +5V
8K X 8
SRAM
74LS373
Vcc
+3V
Backup
100K
.0047uF
Header

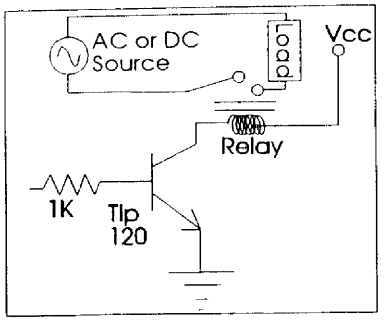
Interface Circuit

9 Pin HEADER

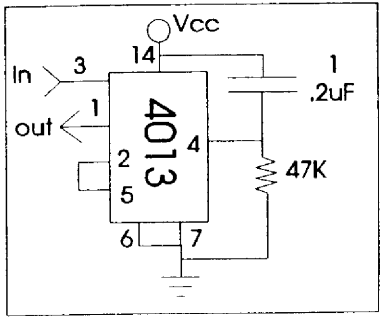




A

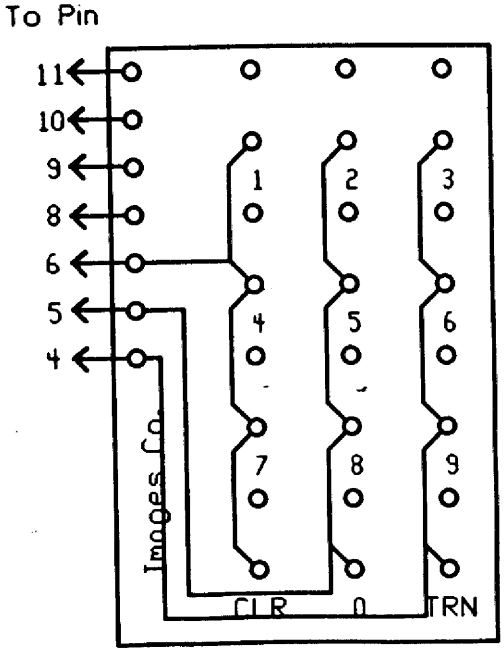


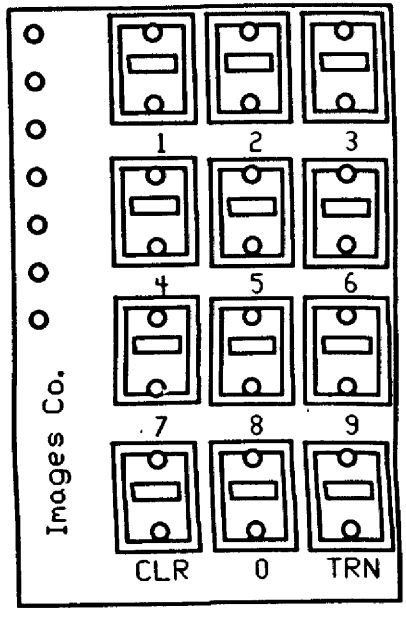
B



C

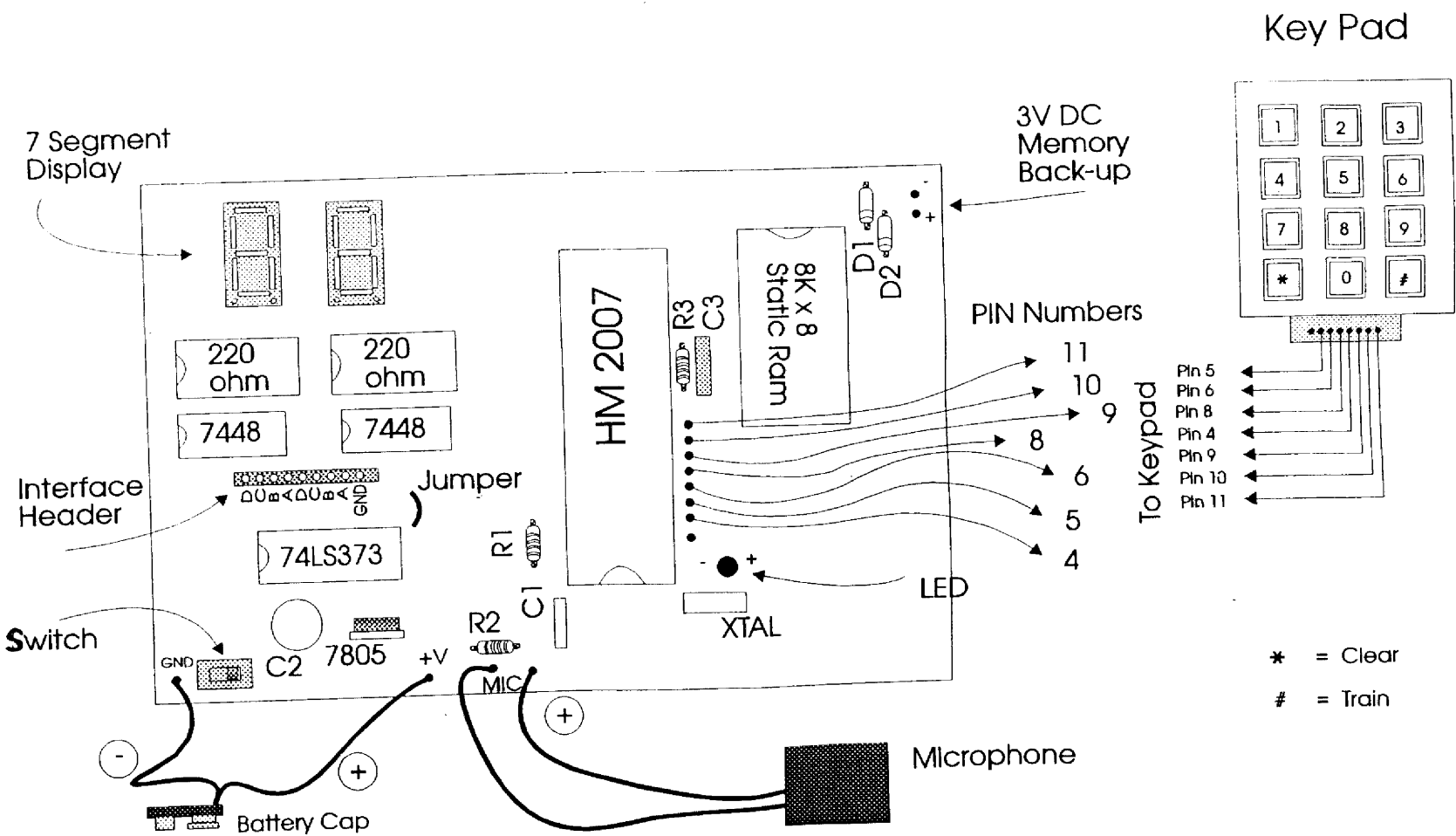
Keypad Assembly





Mount & solder 12 switches to keypad pcb as shown above.

Connect to speech recognition board using pin numbers detailed on the left.


Key Pad
1
2
3
4
5
6
7
8
9
*
0
#
7 Segment Display
3V DC Memory Back-up
220 ohm
220 ohm
7448
7448
HM 2007
R3
C3
8K x 8 Static Ram
D1
D2
PIN Numbers
11
10
9
8
6
5
4
To Keypad
Pin 5
Pin 6
Pin 8
Pin 4
Pin 9
Pin 10
Pin 11
Interface Header
GND
Jumper
74LS373
R1
C1
LED
XTAL
Switch
GND
C2
7805
+V
R2
MIC
* = Clear
# = Train
Microphone
Battery Cap